

Minimax Euclidean Separation Rates for Testing Convex Hypotheses in \mathbb{R}^d

Gilles Blanchard and Alexandra Carpentier
and Maurilio Gutzeit*

*Universität Potsdam, Institut für Mathematik
Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany
e-mail: gilles.blanchard/carpentier/gutzeit@math.uni-potsdam.de*

Abstract: We consider composite-composite testing problems for the expectation in the Gaussian sequence model where the null hypothesis corresponds to a convex subset \mathcal{C} of \mathbb{R}^d . We adopt a minimax point of view and our primary objective is to describe the smallest Euclidean distance between the null and alternative hypotheses such that there is a test with small total error probability. In particular, we focus on the dependence of this distance on the dimension d and the sample size/variance parameter n giving rise to the minimax separation rate. In this paper we discuss lower and upper bounds on this rate for different smooth and non-smooth choices for \mathcal{C} .

Keywords and phrases: composite testing problem, minimax testing.

1. Introduction

In this paper we consider the problem of testing whether a vector $\mu \in \mathbb{R}^d$ belongs to a convex subset \mathcal{C} of \mathbb{R}^d with $d \in \mathbb{N}$, based on a noisy observation X obtained from the Gaussian sequence model with variance scaling parameter $n \in \mathbb{N}$, i.e.

$$X = \mu + \frac{1}{\sqrt{n}}\epsilon, \quad (1.1)$$

where ϵ is a standard Gaussian vector. More precisely, in an L_2 sense, we aim at finding the order of magnitude of the smallest separation distance $\rho > 0$ from \mathcal{C} such that the testing problem

$$H_0 : \mu \in \mathcal{C} \quad \text{vs.} \quad H_\rho : \inf_{c \in \mathcal{C}} \|\mu - c\|_2 \geq \rho, \quad (1.2)$$

where $\|\cdot\|_2$ is the Euclidean distance, can be solved in the following sense: For $\eta \in (0, \frac{1}{2})$, we can construct a uniformly η -consistent test φ for (1.2), i.e.

$$\sup_{\mu \in \mathcal{C}} \mathbb{P}_\mu(\varphi = 1) + \sup_{\mu \in \mathcal{A}_\rho} \mathbb{P}_\mu(\varphi = 0) \leq \eta, \quad (1.3)$$

where $\mathcal{A}_\rho \subset \mathbb{R}^d$ corresponds to H_ρ . We write $\rho^*(\mathcal{C}) := \rho^*(\mathcal{C}, d, n, \eta)$ for the minimax-optimal separation distance ρ for this problem, i.e. the smallest distance ρ that enables the existence of such a test. See Section 2 for a more precise description of the model and relevant quantities.

*Corresponding author

An instance of this problem that was extensively studied is signal detection, i.e. the case where \mathcal{C} is a singleton, see e.g. [11] for an extensive survey of this problem and also [2]. From this literature, we can deduce that the minimax-optimal order of $\rho^*(\mathcal{C})$ in this case is

$$\frac{d^{1/4}}{\sqrt{n}}.$$

In its general form however, this problem is a composite-composite testing problem (i.e. neither \mathcal{C} nor \mathcal{A}_ρ is only a singleton). A versatile way of solving such testing problems was introduced in [9], where the authors combine signal detection ideas with a covering of the null hypothesis, for deriving minimax optimal testing procedures for a composite-composite testing problems, provided that the null hypothesis is not too large (i.e. that its entropy number is not too large, see Assumption (A3) in [9]). In this case, the authors prove that the minimax optimal testing separation rate in this composite-composite testing problem is the same as the signal detection separation rate, namely $\frac{d^{1/4}}{\sqrt{n}}$. This idea can be generalised also in the case where the null hypothesis is “too large” (when Assumption (A3) in [9] is not satisfied), and in this case this approach implies that an upper bound on the minimax rate of separation is the sum of the signal detection rate and the optimal estimation rate in the null hypothesis \mathcal{C} - see [5] for an illustration of this for a specific convex shape. When using this technique, it results that the smaller the entropy of \mathcal{C} , the smaller the separation rate.

This idea has the advantage of generality, but is nevertheless sub-optimal in many simple cases. For instance in the case where the convex \mathcal{C} is a half-space, the minimax-optimal separation rate is $\frac{1}{\sqrt{n}}$, which is much smaller than the minimax-optimal signal detection rate, and that even though a half-space has a much larger entropy (it is even infinite), and larger dimension, than a single point. See Section 3 for an extended discussion on this case. This highlights the fact that for such a testing problem, it is in many cases not the entropy, or size, of the null hypothesis that drives the rate, but rather some other properties of the shape \mathcal{C} .

In order to overcome the limitations of this approach, some other ideas were proposed. A first line of work is in the paper [3], where the authors consider the general testing problem (1.2), but for separation in $\|\cdot\|_2$ -norm instead of $\|\cdot\|_\infty$ -norm. Since any convex can be written as a intersection of half-spaces, they rewrite the problem as a multiple testing problem. This approach is quite fruitful, but the $\|\cdot\|_\infty$ -norm results translate in a non-optimal way in $\|\cdot\|_2$ -norm in high dimension, in terms of the dependence in the dimension d . A second main direction that was investigated was to consider testing for some *specific* convex shapes, as e.g. the cone of positive, monotone, or convex functions, see e.g. [12], or also balls for some metrics [10, 13, 7]. These papers exhibit the minimax-optimal separation distance - or near optimal distance, in some cases of [12] and [13] - for the specific convex shapes that are considered. The models in [12, 13, 7] are different from our model - as they consider functional estimation - and the results are therefore not immediately applicable to our setting.

Besides, they consider specific shapes, i.e. cones in the two first papers, and a smoothness ball in the third paper, but they do not provide results for a more general null hypothesis shape. In Sections 3 and 5, we derive results for our model and shapes related to those of these papers - namely the positive orthant and the euclidian ball - in order to relate our work with these earlier results. Finally, a last type of results that are related to our problem is the case where the null hypothesis can be parametrised, see e.g. [8] where the authors consider shapes that can be parametrised by a quadratic functional. This approach and their results are very suggestive, as they suggest that the smoothness of the shape has an impact on the testing rate.

In this paper, we want to take a more general approach toward the testing problem (1.2). In Section 3, we expose the range of possible separation rates by demonstrating that, without any further assumptions on \mathcal{C} , the statement

$$\frac{1}{\sqrt{n}} \lesssim \rho^*(\mathcal{C}) \lesssim \frac{\sqrt{d}}{\sqrt{n}} \quad (1.4)$$

is sharp up to $\ln(d)$ -factors. After that, in Sections 4 and 5, we investigate the potential of a geometric smoothness property of the surface of \mathcal{C} . Despite its simplicity, this property takes us quite far: In particular, given any separation rate satisfying (1.4), it allows for constructing a set \mathcal{C} exhibiting this rate up to $\ln(d)$ -factors.

2. Setting

Let $d, n \in \mathbb{N}$. We consider the d -dimensional statistical model

$$X = \mu + \frac{1}{\sqrt{n}}\epsilon, \quad (2.1)$$

where $\mu \in \mathbb{R}^d$ is unknown and ϵ is a standard Gaussian vector, written $\epsilon \sim \mathcal{N}(\mathbb{O}_d, \mathbb{I}_d)$ with the null vector \mathbb{O}_d and the identity matrix \mathbb{I}_d . Now, let $\mathcal{C} \subsetneq \mathbb{R}^d$ be non-empty and convex and for $x \in \mathbb{R}^d$,

$$\text{dist}(x, \mathcal{C}) := \inf_{c \in \mathcal{C}} \|x - c\|_2,$$

where $\|\cdot\|_2$ denotes Euclidean distance, i.e. $\|x\|_2 = \sqrt{\sum_{i=1}^d x_i^2}$. A corresponding open Euclidean ball with center $z \in \mathbb{R}^d$ and radius $r > 0$ is denoted $B(z, r)$.

For $\rho > 0$, we are interested in the testing problem

$$H_0 : \mu \in \mathcal{C} \quad \text{vs.} \quad H_\rho : \text{dist}(\mu, \mathcal{C}) \geq \rho \quad (2.2)$$

and we write $\mathcal{A}_\rho := \{x \in \mathbb{R}^d \mid \text{dist}(x, \mathcal{C}) \geq \rho\}$.

Our aim is to find the smallest value of ρ such that testing (2.2) with prescribed total error probability is possible in a minimax sense, i.e. the quantity

$$\begin{aligned}\rho^*(\mathcal{C}) &:= \rho^*(\mathcal{C}, d, n, \eta) \\ &= \inf\{\rho > 0 \mid \exists \text{ test } \varphi : \sup_{\mu \in \mathcal{C}} \mathbb{P}_\mu(\varphi = 1) + \sup_{\mu \in \mathcal{A}_\rho} \mathbb{P}_\mu(\varphi = 0) \leq \eta\}\end{aligned}$$

for $\eta \in (0, \frac{1}{2})$. Here, a test φ is a measurable function $\varphi : \mathbb{R}^d \rightarrow \{0, 1\}$.

In particular, we focus on the dependence of $\rho^*(\mathcal{C})$ on the dimension d and sample size/ variance parameter n . In terms of notation, this is done by using the symbols \lesssim, \gtrsim and \approx as follows: We write $\rho^*(\mathcal{C}) \lesssim g_{\mathcal{C}}(d, n)$ where $g_{\mathcal{C}}(d, n)$ depends only on \mathcal{C}, n, d , if for some $w(\eta) > 0$ depending only on η and $C > 0$ a universal constant, $\rho^*(\mathcal{C}) \leq Cw(\eta)g_{\mathcal{C}}(d, n)$. We define in a similar way the symbol \gtrsim (other direction).

Now, if $g_{\mathcal{C}}(d, n) \lesssim \rho^*(\mathcal{C}) \lesssim g_{\mathcal{C}}(d, n)$, we write $\rho^*(\mathcal{C}) \approx g_{\mathcal{C}}(d, n)$ and call $g_{\mathcal{C}}(d, n)$ the *minimax Euclidean separation rate* for (2.2) or simply *separation rate*.

Remark 2.1. In the proofs for upper bounds on $\rho^*(\mathcal{C})$ it is necessary to consider the type- I and type- II errors $\sup_{\mu \in \mathcal{C}} \mathbb{P}_\mu(\varphi = 1)$ and $\sup_{\mu \in \mathcal{A}_\rho} \mathbb{P}_\mu(\varphi = 0)$ separately

leading to parameters $\alpha, \beta \in (0, \frac{1}{2})$ rather than η . However, this does not affect the separation rate. For the sake of consistency in notation, we will state the exact constants $w(\eta)$ in upper bounds with $\alpha = \beta = \frac{\eta}{2}$. In these statements and in the proofs, we use the abbreviation $v_x := \ln(\frac{1}{x})$, $x > 0$.

3. A General Guarantee and Extreme Cases

The rate $\rho^*(\mathcal{C})$ clearly depends on \mathcal{C} .

Let us firstly examine a simple, essentially one-dimensional case, namely a half-space.

Theorem 3.1. *Let $\mathcal{C} = \mathcal{C}_{\text{HS}} = (-\infty, 0] \times \mathbb{R}^{d-1}$ (if $d = 1$, $\mathcal{C}_{\text{HS}} = (-\infty, 0]$). Then, in the testing problem (2.2), we have*

$$\sqrt{\frac{1}{n} \ln(1 + 4(1 - \eta)^2)} \leq \rho^*(\mathcal{C}) \leq \sqrt{\frac{8}{n} v_\eta}$$

and therefore

$$\rho^*(\mathcal{C}) \approx \frac{1}{\sqrt{n}}.$$

Remark 3.2. Note that the lower bound in the previous theorem is valid for any choice of convex \mathcal{C} such that \mathcal{C} and $\mathbb{R}^d \setminus \mathcal{C}$ are non-empty:

$$\frac{1}{\sqrt{n}} \lesssim \rho^*(\mathcal{C}).$$

This is easily verified using the fact that due to convexity there is a hyperplane in \mathbb{R}^d tangent to \mathcal{C} such that one of the two corresponding half-spaces fully contains \mathcal{C} . In the case of Theorem 3.1, the hyperplane $\{0\} \times \mathbb{R}^{d-1}$ has this role.

Now, on the other hand, making no additional assumptions about \mathcal{C} , a natural choice φ for solving (2.2) is a plug-in test based on confidence balls. This gives rise to the following general upper bound:

Theorem 3.3. *Let \mathcal{C} be an arbitrary convex subset of \mathbb{R}^d such that \mathcal{C} and $\mathbb{R}^d \setminus \mathcal{C}$ are non-empty. Then, in the testing problem (2.2), we have*

$$\rho^*(\mathcal{C}) \leq 2\sqrt{\frac{d}{n} + \frac{2}{n}\sqrt{dv_{\eta/2}} + \frac{2}{n}v_{\eta/2}}$$

and therefore

$$\rho^*(\mathcal{C}) \lesssim \frac{\sqrt{d}}{\sqrt{n}}.$$

Remark 3.4. Note that this upper bound is the rate of estimation of μ in L_2 norm in the model (2.1) (See Equation (6.2) in Section 6.1.2).

Remark 3.5. From Remark 3.2 and Theorem 3.3, it is clear that for any \mathcal{C} being a convex subset of \mathbb{R}^d such that \mathcal{C} and $\mathbb{R}^d \setminus \mathcal{C}$ are non-empty, it holds that

$$\frac{1}{\sqrt{n}} \lesssim \rho^*(\mathcal{C}) \lesssim \frac{\sqrt{d}}{\sqrt{n}}.$$

Given Remarks 3.2 and 3.5, it is natural to ask if the upper bound in Theorem 3.3 is also sharp, in the sense that there is a choice of \mathcal{C} that requires the separation rate $\frac{\sqrt{d}}{\sqrt{n}}$, at least up to logarithmic factors. It turns out that the answer is yes when \mathcal{C} is taken to have the shape of an orthant:

Theorem 3.6. *Let $\mathcal{C} = \mathcal{C}_O = (-\infty, 0]^d$, $d \geq 2$. Then, if $\eta \in (0, \frac{1}{6})$, for the testing problem (2.2), we have*

$$\rho^*(\mathcal{C}) \geq \frac{\sqrt{d}}{288(\lceil \ln(d) \rceil + 1/3)^2 \sqrt{n}}$$

and therefore

$$\frac{\sqrt{d}}{\ln(d)^2 \sqrt{n}} \lesssim \rho^*(\mathcal{C}) \lesssim \frac{\sqrt{d}}{\sqrt{n}}.$$

A version of this result was already proven in [12]; however, the influence of n is not considered there (which appears to be due to the fact that the result only serves as a lemma for another theorem), $\eta = \frac{1}{8}$ is fixed and the bound holds for $d \geq D$, where $D \in \mathbb{N}$ is fixed but not explicit. We provide a partly alternative or adapted proof in Section 6.2.3 that leads to the statement above.

4. A Simple Smoothness Property

Clearly, the two extreme cases \mathcal{C}_{HS} and \mathcal{C}_{O} differ significantly with respect to smoothness of their surfaces. Based on this observation, in order to be able to handle $\rho^*(\mathcal{C})$ more flexibly, we propose to describe convex sets by their surfaces' degree of smoothness, where the surface of a set $S \in \mathbb{R}^d$ is denoted by ∂S . To begin with, we examine the potential of the following very simple and purely geometric smoothness concept:

Definition 4.1. Let $R > 0$ and $S \subseteq \mathbb{R}^d$ be nonempty. S is called **R -smooth** if

$$\forall x \in \partial S \exists z \in S : x \in \overline{B(z, R)} \subseteq \overline{S}. \quad (4.1)$$

Remark 4.2. The definition of R -smoothness is closely related to the so-called R -rolling condition employed in [1]. In fact, R -smoothness of S is equivalent to saying that $\mathbb{R}^d \setminus S$ fulfils the R -rolling condition.

Another related concept worth mentioning is the radius of curvature, though the connection is more subtle: The radius of curvature at a point $x \in \partial S$ would be the largest radius R such that there is a ball $B(z, R)$ tangent to ∂S at x such that $z - x$ points into \mathcal{C} . Hence, it is possible that the infimum of these radii with respect to $x \in \partial S$ corresponds to the parameter R in our previous definition. However, note that the condition $\overline{B(z, R)} \subseteq \overline{S}$ is then still not guaranteed in general, so that the indicated notion of a uniform radius of curvature is not equivalent to R -smoothness.

Since smoothness is usually defined as a property of a function, we provide a suggestion for how to cast the above concept in that context for a convex set \mathcal{C} : Given any $x \in \partial \mathcal{C}$, without loss of generality (w.l.o.g.) apply a mapping G that is orthogonal (represented by an orthogonal matrix) up to translation such that $x' = G(x) = \mathbb{O}_d$ and $\mathcal{C}' := G(\mathcal{C}) \subseteq \mathbb{R}^{d-1} \times [0, \infty)$. Now, let f be the function describing $\partial \mathcal{C}'$ over $B := B(\mathbb{O}_{d-1}, R)$. The following lemma states sufficient conditions for R -smoothness at $x \in \mathcal{C}$:

Lemma 4.3. In the situation described above, if f is twice differentiable on B (i.e. the gradient $\nabla f(\cdot)$ and Hessian matrix $Hf(\cdot)$ exist), the following conditions are sufficient in order that $\partial \mathcal{C}'$ and $B((\mathbb{O}_{d-1}, R), R)$ (the ball associated to x in (4.1)) do not intersect in $B \times (0, R]$:

$$\begin{cases} \nabla f(\mathbb{O}_{d-1}) = \mathbb{O}_{d-1}, \\ \forall x \in B \setminus \{\mathbb{O}_{d-1}\} : 0 \leq \lambda_{\min}(Hf(x)), \lambda_{\max}(Hf(x)) \leq \frac{1}{R}, \end{cases}$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are the lowest and highest eigenvalues of a real symmetric matrix, respectively.

Now, let us examine how the additional assumption of R -smoothness may affect the general upper bound of Theorem 3.3:

Theorem 4.4. *If \mathcal{C} is R -smooth for $R > 0$, for the testing problem (2.2), we have*

$$\rho^*(\mathcal{C}) \leq \sqrt{\frac{2}{n}v_{\eta/8} + \frac{d}{2nR}} + \frac{2}{nR}\sqrt{dv_{\eta/4}} + \frac{1}{nR}v_{\eta/4} + \sqrt{\frac{2}{n}v_{\eta/2}}$$

and therefore, taking Theorem 3.3 into account,

$$\rho^*(\mathcal{C}) \lesssim \max\left(\frac{1}{\sqrt{n}}, \min\left(\frac{\sqrt{d}}{\sqrt{n}}, \frac{d}{nR}\right)\right)$$

The following result confirms that this upper bound can be sharp up to $\ln(d)$ factors, namely in the case where \mathcal{C} is taken as the R -inflated negative orthant:

Theorem 4.5. *Let $d \geq 3$, $\eta \in (0, \frac{1}{6})$ and*

$$\mathcal{C} = \mathcal{C}_{\text{IO}} = \mathcal{C}_{\text{O}} + B(\text{O}_d, R) = \bigcup_{z \in \mathcal{C}_{\text{O}}} B(z, R).$$

Then, in the testing problem (2.2), we have with $s = \frac{1}{144(\lceil \ln(d) \rceil + 1/3)^2 \sqrt{n}}$

$$\rho^*(\mathcal{C}) \geq \frac{(d-1)s^2}{8R + 4\sqrt{d-1}s} \geq \frac{1}{16} \min\left(\frac{(d-1)s^2}{2R}, \sqrt{d-1}s\right)$$

and therefore,

$$\rho^*(\mathcal{C}) \gtrsim \max\left(\frac{1}{\sqrt{n}}, \min\left(\frac{1}{\ln(d)^4} \cdot \frac{d}{nR}, \frac{1}{\ln(d)^2} \cdot \frac{\sqrt{d}}{\sqrt{n}}\right)\right).$$

5. Discussion

The concept of R -smoothness allows for the construction of hypotheses \mathcal{C} with any separation rate $\frac{1}{\sqrt{n}} \lesssim \rho^*(\mathcal{C}) \lesssim \frac{\sqrt{d}}{\sqrt{n}}$, up to $\ln(d)$ -factors. On the other hand, we must acknowledge that R -smoothness is too weak a concept to fully describe the difficulty of testing an arbitrary \mathcal{C} ; an examination of the natural R -smooth set, namely a ball of radius R , provides clear evidence of this drawback. The result is a direct generalisation of the known rate $\rho^*(\mathcal{C}) \sim \frac{d^{\frac{1}{4}}}{\sqrt{n}}$ in the signal detection setting, see [2].

Theorem 5.1. *Let $d \geq 3$. If $\mathcal{C} = \mathcal{C}_{\text{B}} = B(z, R)$, $z \in \mathbb{R}^d$ and $R > 0$, for the testing problem (2.2), we have for $s := \frac{\sqrt{d-1}}{n} \sqrt{\frac{2}{e} \ln(1 + 4(1-\eta)^2)}$*

$$\rho^*(\mathcal{C}) \geq \frac{s}{2\sqrt{s+R^2}} \gtrsim \min\left(\frac{d^{\frac{1}{4}}}{\sqrt{n}}, \frac{\sqrt{d}}{nR}\right)$$

and also

$$\begin{aligned} \rho^*(\mathcal{C}) &\leq \min\left(2\sqrt{2}\frac{d^{\frac{1}{4}}}{n^{\frac{1}{2}}}\sqrt{v_{\eta/2}} + 3\sqrt{\frac{2}{n}v_{\eta/2}}, \frac{2\sqrt{d}}{nR + 2\sqrt{nv_{\eta/2}}}\sqrt{v_{\eta/2}} + 3\sqrt{\frac{2}{n}v_{\eta/2}}\right) \\ &\lesssim \max\left(\frac{1}{\sqrt{n}}, \min\left(\frac{d^{\frac{1}{4}}}{\sqrt{n}}, \frac{\sqrt{d}}{nR}\right)\right). \end{aligned}$$

Therefore,

$$\rho^*(\mathcal{C}) \asymp \max \left(\frac{1}{\sqrt{n}}, \min \left(\frac{d^{\frac{1}{4}}}{\sqrt{n}}, \frac{\sqrt{d}}{nR} \right) \right). \quad (5.1)$$

Clearly, Theorem 4.5 does not capture this case. As a consequence, future work will be concerned with finding a stronger concept, possibly a localised version of R -smoothness, that ideally allows for describing $\rho^*(\mathcal{C})$ for any choice of \mathcal{C} . However, we suspect this to be quite an ambitious goal.

6. Proofs

6.1. General Preparations

6.1.1. Techniques for Obtaining Lower Bounds

We employ a classical Bayesian approach for proving lower bounds, see references in [2] for its origins. We briefly give the main theoretical ingredients of this approach for our setting:

Let ν_0 be a distribution with $S_0 := \text{supp}(\nu_0) \subseteq \mathcal{C}$ and ν_ρ be a distribution with $S_\rho := \text{supp}(\nu_\rho) \subseteq \{\mu \in \mathbb{R}^d \mid \text{dist}(\mu, \mathcal{C}) = \rho\}$ (priors). For instance, Dirac priors on some $x \in \mathbb{R}$ will be denoted δ_x . Furthermore, for $i \in \{0, \rho\}$, let \mathbb{P}_{ν_i} be the resulting distribution of X given $\mu \sim \nu_i$. Now, we see that for any test $\varphi = \mathbb{1}_A$, $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} \sup_{\mu \in \mathcal{C}} \mathbb{P}_\mu(\varphi = 1) + \sup_{\mu \in \mathcal{A}_\rho} \mathbb{P}_\mu(\varphi = 0) &\geq \mathbb{P}_{\nu_0}(\varphi = 1) + \mathbb{P}_{\nu_\rho}(\varphi = 0) \\ &= 1 - (\mathbb{P}_{\nu_\rho}(A) - \mathbb{P}_{\nu_0}(A)) \\ &\geq 1 - \sup_A |\mathbb{P}_{\nu_\rho}(A) - \mathbb{P}_{\nu_0}(A)| \\ &= 1 - \frac{1}{2} \|\mathbb{P}_{\nu_\rho} - \mathbb{P}_{\nu_0}\|_{\text{TV}} \\ &\geq 1 - \frac{1}{2} \left(\int_{\mathbb{R}^d} \left(\frac{d\mathbb{P}_{\nu_\rho}}{d\mathbb{P}_{\nu_0}} \right)^2 d\mathbb{P}_{\nu_0} - 1 \right)^{\frac{1}{2}}. \end{aligned}$$

This justifies the following reasoning used for each lower bound proof in the present paper:

For any $\tilde{\rho} > 0$ such that either

$$\frac{1}{2} \|\mathbb{P}_{\nu_0} - \mathbb{P}_{\nu_{\tilde{\rho}}}\|_{\text{TV}} \leq 1 - \eta \in \left(\frac{1}{2}, 1 \right)$$

or

$$\int_{\mathbb{R}^d} \left(\frac{d\mathbb{P}_{\nu_\rho}}{d\mathbb{P}_{\nu_0}} \right)^2 d\mathbb{P}_{\nu_0} \leq 1 + 4(1 - \eta)^2 \in (2, 5), \quad (6.1)$$

for the testing problem (2.2), we have

$$\rho^*(\mathcal{C}) \geq \tilde{\rho}.$$

6.1.2. Concentration Properties of Gaussian and χ^2 Random Variables

We will repeatedly make use of the following classical properties of $N \sim \mathcal{N}(0, \sigma^2)$ and $Z \sim \chi_\lambda^2(d)$ (that is, a χ^2 -distribution with d degrees of freedom and non-centrality parameter $\lambda \geq 0$): For any $\delta \in (0, 1)$, the following concentration inequalities hold:

$$\begin{aligned} \text{(I)} \quad & \mathbb{P}(N \geq \sigma\sqrt{2v_\delta}) \leq \delta \\ \text{(II)} \quad & \mathbb{P}(Z \geq d + \lambda + 2\sqrt{(d+2\lambda)v_\delta + 2v_\delta}) \leq \delta, \\ \text{(III)} \quad & \mathbb{P}(Z \leq d + \lambda - 2\sqrt{(d+2\lambda)v_\delta}) \leq \delta, \end{aligned} \quad (6.2)$$

See [4] for proofs of (6.2.II) and (6.2.III).

6.1.3. Lower Bounds for Specific $\sqrt{\cdot}$ -Functions

The following bounds are easily obtained via Taylor expansion; however, we will employ them on several occasions which makes it convenient to mention them here.

For any $a > 0, b \in \mathbb{R}$, we have

$$\frac{a}{2\sqrt{a+b^2}} \leq \sqrt{a+b^2} - b \leq \frac{a}{2b} \quad (6.3)$$

and for any $b > 0, a \in (0, b^2)$ we have

$$b - \sqrt{b^2 - a} \geq \frac{a}{2b} \quad (6.4)$$

6.2. Proofs for Section 3

6.2.1. Proof of Theorem 3.1

Proof. We prove independently that $\rho^*(\mathcal{C})$ is lower and upper bounded in order by $\frac{1}{\sqrt{n}}$.

Lower Bound. In accordance with the framework in Section 6.1.1, we verify that the bound holds in the special case $\nu_0 = \delta_{\mathbf{O}_d}$ and $\nu_\rho = \delta_{\rho \cdot \mathbf{e}_1}$. Since both the null and alternative hypotheses are simple, the corresponding density functions $F_{\nu_0}(x)$ and $F_{\nu_\rho}(x)$ are readily given and we obtain

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{F_{\nu_\rho}^2}{F_{\nu_0}}(x) \, dx &= \sqrt{\frac{n}{2\pi}}^d \int_{\mathbb{R}^d} \exp\left(-n(x_1 - \rho)^2 + \frac{n}{2}x_1^2\right) \\ &\quad \cdot \exp\left(-\frac{n}{2}(x_2^2 + x_3^2 + \dots + x_d^2)\right) \, dx \\ &= \sqrt{\frac{n}{2\pi}} \int_{\mathbb{R}} \exp\left(-n(x_1 - \rho)^2 + \frac{n}{2}x_1^2\right) \, dx_1 \\ &= \sqrt{\frac{n}{2\pi}} \exp(n\rho^2) \int_{\mathbb{R}} \exp\left(-\frac{n}{2}(x_1 - 2\rho)^2\right) \, dx_1 \\ &= \exp(n\rho^2). \end{aligned}$$

Therefore inequality (6.1) is satisfied (with equality) when the latter quantity is equal to $1 + 4(1 - \eta)^2$, i.e. equivalently for

$$\rho = \sqrt{\frac{1}{n} \ln(1 + 4(1 - \eta)^2)}.$$

This yields the claim.

Upper Bound. Given $\alpha, \beta \in (0, \frac{1}{2})$, let $\delta = \min(\alpha, \beta)$ and $\tau_\delta = \sqrt{\frac{2}{n} v_\delta}$. Define the test

$$\varphi(X) = \mathbb{1}_{\{X_1 \geq \tau_\delta\}}.$$

Then for any $\mu \in \mathcal{C}$, we have

$$\mathbb{P}_\mu(\varphi(X) = 1) \leq \mathbb{P}\left(\frac{1}{\sqrt{n}}\epsilon_1 \geq \tau_\delta\right) \stackrel{(6.2.I)}{\leq} \delta \leq \alpha.$$

On the other hand, let now $\rho = 2\tau_\delta$. Then for any $\mu \in \mathcal{A}_\rho$, we have

$$\mathbb{P}_\mu(\varphi(X) = 0) \leq \mathbb{P}\left(2\tau_\delta + \frac{1}{\sqrt{n}}\epsilon_1 \leq \tau_\delta\right) \leq \mathbb{P}\left(\frac{1}{\sqrt{n}}\epsilon_1 \leq -\tau_\delta\right) \stackrel{(6.2.I)}{\leq} \delta \leq \beta.$$

This concludes the proof since $\rho \approx \frac{1}{\sqrt{n}}$. \square

6.2.2. Proof of Theorem 3.3

Proof. Given $\alpha, \beta \in (0, \frac{1}{2})$, let $\delta = \min(\alpha, \beta)$ and $\tau_\delta = \frac{d}{n} + \frac{2}{n}\sqrt{dv_\delta} + \frac{2}{n}v_\delta$. Define the test

$$\varphi(X) = \mathbb{1}_{\{B(X, \sqrt{\tau_\delta}) \cap \mathcal{C} = \emptyset\}} = \mathbb{1}_{\{\text{dist}(X, \mathcal{C}) \geq \sqrt{\tau_\delta}\}}.$$

Then for any $\mu \in \mathcal{C}$, we have

$$\mathbb{P}_\mu(\varphi(X) = 1) \leq \mathbb{P}(\|X - \mu\| \geq \sqrt{\tau_\delta}) \leq \mathbb{P}\left(\frac{1}{n}\|\epsilon\|^2 \geq \tau_\delta\right) \stackrel{(6.2.II)}{\leq} \delta \leq \alpha.$$

On the other hand, let now $\rho = 2\sqrt{\tau_\delta}$. Then for any $\mu \in \mathcal{A}_\rho$, we have since $\text{dist}(X, \mathcal{C}) \geq \text{dist}(\mu, \mathcal{C}) - \|X - \mu\|$:

$$\mathbb{P}_\mu(\varphi(X) = 0) \leq \mathbb{P}\left(\frac{1}{n}\|\epsilon\|^2 \geq \tau_\delta\right) \stackrel{(6.2.II)}{\leq} \delta \leq \beta.$$

This concludes the proof since $\sqrt{\tau_\delta} \approx \sqrt{\frac{d}{n}}$. \square

6.2.3. Proof of Theorem 3.6

Proof. The arguments of this proof are related to the ones used in [12] and [6]. We decompose the proof into several steps.

1. Choice of priors.

We make use of the following lemma used and explained in [12] :

Lemma. For any $M \in \mathbb{N}$ and $b > 0$, there are distributions $\tilde{\nu}_0$ and $\tilde{\nu}_1$ with the following properties:

$$\begin{aligned} \text{(I)} \quad & \text{supp}(\tilde{\nu}_0) \subseteq [-b, 0], \quad \text{supp}(\tilde{\nu}_1) \subseteq [-b, 0] \cup \left\{ \frac{b}{4M^2} \right\} \\ \text{(II)} \quad & \tilde{\nu}_1 \left(\left\{ \frac{b}{4M^2} \right\} \right) \geq \frac{1}{2} \\ \text{(III)} \quad & \forall k \in \{0, 1, \dots, M\} : \int x^k \tilde{\nu}_0(dx) = \int x^k \tilde{\nu}_1(dx). \end{aligned} \tag{6.5}$$

For now, let $\tilde{\nu}_i$ be such distributions and $\nu_i = \tilde{\nu}_i^{\otimes d}$, $i \in \{0, 1\}$; M, b and ρ

will be specified later. Furthermore, writing $\sigma^2 = \frac{1}{n}$, let

$$\mathbb{P}_i = (\tilde{\nu}_i * \mathcal{N}(0, \sigma^2))^{\otimes d}, \quad i \in \{0, 1\},$$

where $*$ denotes convolution. Clearly, the corresponding density function can be written as

$$F_i(x) = \prod_{j=1}^d (\mathbb{E}_{\mu_j \sim \tilde{\nu}_i} [\phi(x_j; \mu_j, \sigma^2)]), \quad i \in \{0, 1\}.$$

where $\phi(x; \mu, \sigma^2)$ is the density of $\mathcal{N}(\mu, \sigma^2)$. It will be convenient to examine the case $d = 1$, denoted by $\tilde{\mathbb{P}}_i$.

Note that \mathbb{P}_0 is in accordance with \mathbb{P}_{ν_0} from Section 6.1.1, but the construction of ν_1 does not warrant the notion of Euclidean distance we are interested in, hence the slight difference in notation. This technical obstacle is necessary for the property (6.5.III), but it can be resolved for a small price, which we explain in the last step of this proof.

2. Controlling the total variation distance.

Based on our construction and writing $u := \frac{b}{4M^2}$, we have for $i \in \{0, 1\}$

$$\begin{aligned} \mathbb{E}_{\mu \sim \tilde{\nu}_i} [\phi(x; \mu, \sigma^2)] &= \phi(x; 0, \sigma^2) \int \exp\left(\frac{2x\mu - \mu^2}{2\sigma^2}\right) \tilde{\nu}_i(d\mu) \\ &= \phi(x; 0, \sigma^2) \int \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{2x\mu - \mu^2}{2\sigma^2}\right)^k \tilde{\nu}_i(d\mu) \\ &= \phi(x; 0, \sigma^2) \int \sum_{k=0}^{\infty} \frac{1}{k! (2\sigma^2)^k} (2x\mu - \mu^2)^k \tilde{\nu}_i(d\mu) \\ &= \phi(x; 0, \sigma^2) \sum_{k=0}^{\infty} \sum_{j=0}^k \binom{k}{j} \frac{(-1)^{k-j} 2^j}{k! (2\sigma^2)^k} x^j \int \mu^{2k-j} \tilde{\nu}_i(d\mu). \end{aligned}$$

As a consequence, by (6.5.III), this yields

$$\begin{aligned}
\|\tilde{\mathbb{P}}_1 - \tilde{\mathbb{P}}_0\|_{\text{TV}} &= \int |\mathbb{E}_{\mu \sim \tilde{\nu}_1}[\phi(x; \mu, \sigma^2)] - \mathbb{E}_{\mu \sim \tilde{\nu}_0}[\phi(x; \mu, \sigma^2)]| \, dx \\
&= \int \left| \sum_{k=\lfloor M/2 \rfloor + 1}^{\infty} \sum_{j=0}^k \binom{k}{j} \frac{(-1)^{k-j} 2^j}{k! (2\sigma^2)^k} x^j \right. \\
&\quad \cdot \left[\int \mu^{2k-j} \tilde{\nu}_1(d\mu) - \int \mu^{2k-j} \tilde{\nu}_0(d\mu) \right] \phi(x; 0, \sigma^2) \, dx \\
&\leq \sum_{k=\lfloor M/2 \rfloor + 1}^{\infty} \sum_{j=0}^k \binom{k}{j} \frac{2^j}{k! (2\sigma^2)^k} \underbrace{\left| \int \mu^{2k-j} \tilde{\nu}_1(d\mu) - \int \mu^{2k-j} \tilde{\nu}_0(d\mu) \right|}_{=I_{2k-j}} \\
&\quad \cdot \int |x|^j \phi(x, 0, \sigma^2) \, dx.
\end{aligned}$$

We now use the bound $I_{2k-j} \leq 2b^{2k-j}$ and the fact that

$$\int_{\mathbb{R}} |x|^j \phi(x, 0, \sigma^2) \, dx = \frac{\sigma^j \sqrt{2}^j}{\sqrt{\pi}} \Gamma(j/2 + 1) \leq \frac{\sigma^j \sqrt{2}^j k!}{\sqrt{\pi}}$$

(see [14]) to obtain

$$\begin{aligned}
\|\tilde{\mathbb{P}}_1 - \tilde{\mathbb{P}}_0\|_{\text{TV}} &\leq \frac{2}{\sqrt{\pi}} \sum_{k=\lfloor M/2 \rfloor + 1}^{\infty} \sum_{j=0}^k \binom{k}{j} \frac{2^j}{k! (2\sigma^2)^k} b^{2k-j} \sigma^j \sqrt{2}^j k! \\
&= \frac{2}{\sqrt{\pi}} \sum_{k=\lfloor M/2 \rfloor + 1}^{\infty} \left(\frac{b}{2\sigma} \right)^k \sum_{j=0}^k \binom{k}{j} (2\sqrt{2})^j \left(\frac{b}{\sigma} \right)^{k-j} \\
&= \frac{2}{\sqrt{\pi}} \sum_{k=\lfloor M/2 \rfloor + 1}^{\infty} \left[\frac{s}{2} (2\sqrt{2} + s) \right]^k, \quad s := \frac{b}{\sigma}.
\end{aligned}$$

Now, if $t := \frac{s}{2} (2\sqrt{2} + s) \in (0, \frac{1}{2}]$, i.e. $b \in (0, \sigma(\sqrt{3} - \sqrt{2})]$, this yields using $1 - t \geq \frac{1}{2}$:

$$\|\mathbb{P}_1 - \mathbb{P}_0\|_{\text{TV}} \leq d \|\tilde{\mathbb{P}}_1 - \tilde{\mathbb{P}}_0\|_{\text{TV}} \leq \frac{2d}{\sqrt{\pi}} \left[\frac{b}{2\sigma} \left(2\sqrt{2} + \frac{b}{\sigma} \right) \right]^{\lfloor M/2 \rfloor + 1} \leq \frac{2d}{\sqrt{\pi}} 2^{-\lfloor M/2 \rfloor}.$$

Finally, let $\eta' \in (0, \frac{1}{2})$ and choose $b = \frac{1}{4}\sigma = \frac{1}{4\sqrt{n}}$. By straightforward computation we can show that

$$M \geq \frac{2}{\ln(2)} \ln(d) - \frac{\ln(\pi(1 - \eta')^2)}{\ln(2)} \implies \frac{1}{2} \|\mathbb{P}_1 - \mathbb{P}_0\|_{\text{TV}} \leq 1 - \eta'.$$

Here, since $\eta' < \frac{1}{2}$, less than $\frac{1}{2}$ is subtracted; together with $\frac{2}{\ln(2)} \leq 3$ we can therefore fix the solution

$$M = 3\lceil \ln(d) \rceil + 1.$$

3. Application.

Note that this upper bound $\frac{1}{2}\|\mathbb{P}_1 - \mathbb{P}_0\|_{\text{TV}} \leq 1 - \eta'$ does not formally allow for determining a lower bound on $\rho^*(\mathcal{C})$ yet since H_0 and H_1 are not separated in a Euclidean sense. In a final step, we will resolve this by a suitable restriction of H_1 .

Let $Y = \sum_{i=1}^d \mathbb{1}_{\{\mu_i=u\}}$, i.e. the number of coordinates of μ taking the value u . Obviously, if $\mu \sim \nu_1$, we have $Y \sim \text{Bin}(d, \tilde{\nu}_1(\{u\}))$. By property (6.5.II), this yields that if $d \geq 2$,

$$\mathbb{P}_{\mu \sim \nu_1} \left(Y \geq \frac{d}{4} \right) \geq \frac{3}{4}.$$

Now, let $\xi = \{Y \geq \frac{d}{4}\}$ and

$$H'_1 : \mu \sim \nu_1 | \xi.$$

Assuming that for some test φ the relation

$$\mathbb{P}_{\mu \sim \nu_0}(\varphi = 1) + \mathbb{P}_{\mu \sim \nu_1}(\varphi = 0) \geq \eta'$$

holds, we can conclude

$$\begin{aligned} \mathbb{P}_{\nu_0}(\varphi = 1) + \mathbb{P}_{\nu_1|\xi}(\varphi = 0) &= \mathbb{P}_{\nu_0}(\varphi = 1) + 1 - \frac{\mathbb{P}_{\nu_1}(\{\varphi = 1\} \cap \xi)}{\mathbb{P}(\xi)} \\ &\geq \mathbb{P}_{\nu_0}(\varphi = 1) + 1 - \frac{4}{3}\mathbb{P}_{\nu_1}(\varphi = 1) \\ &\geq \mathbb{P}_{\nu_0}(\varphi = 1) + \frac{4}{3}\mathbb{P}_{\nu_1}(\varphi = 0) - \frac{1}{3} \\ &\geq \mathbb{P}_{\nu_0}(\varphi = 1) + \mathbb{P}_{\nu_1}(\varphi = 0) - \frac{1}{3} \\ &\geq \eta' - \frac{1}{3}. \end{aligned}$$

Hence, inference from the testing problem discussed in Steps 1 and 2 to the problem

$$H_0 : \mu \sim \nu_0 \text{ vs. } H'_1 : \mu \sim \nu_1 | \xi$$

is valid as long as $\eta \in (0, \frac{1}{6})$ (η corresponds to $\eta' - \frac{1}{3}$ above).

The following observation concludes the proof: Clearly, H'_1 agrees with \mathcal{A}_ρ for

$$\rho = \frac{\sqrt{d}}{2}u = \frac{\sqrt{d}\sigma}{32M^2} = \frac{\sqrt{d}}{288(\lceil \ln(d) \rceil + 1/3)^2\sqrt{n}} \sim \frac{\sqrt{d}}{\ln(d)^2\sqrt{n}}.$$

□

6.3. Proofs for Section 4

6.3.1. Proof of Lemma 4.3

Proof. Let $x \in B \setminus \{0_{d-1}\}$. By construction, we need to ensure that

$$0 \leq f(x) \leq R - \sqrt{R^2 - \|x\|^2}.$$

Applying Taylor's theorem with Lagrange's remainder yields

$$\exists s \in (0, 1) : f(x) = \frac{1}{2} x^T Hf(sx) x.$$

Clearly, for $f \geq 0$ on B , it is sufficient to require $\lambda_{\min}(Hf(y)) \geq 0$ for $y \in B \setminus \{0_{d-1}\}$. On the other hand, we can use a classical eigenvalue representation to obtain the desired upper bound: For some $s \in (0, 1)$,

$$\begin{aligned} f(x) &= \frac{1}{2} \|x\|^2 \left(\frac{x}{\|x\|} \right)^T Hf(sx) \left(\frac{x}{\|x\|} \right) \\ &\leq \frac{1}{2} \|x\|^2 \max_{\|y\|=1} y^T Hf(sx) y \\ &= \frac{1}{2} \|x\|^2 \lambda_{\max}(Hf(sx)). \end{aligned}$$

This already implies a sufficient condition on Hf for the upper bound:

$$\forall x \in B \setminus \{0_{d-1}\} \forall s \in (0, 1) : \lambda_{\max}(Hf(sx)) \leq \frac{2(R - \sqrt{R^2 - \|x\|^2})}{\|x\|^2}.$$

However, (6.4) tells us that for $z \in (0, R^2)$,

$$R - \sqrt{R^2 - z} \geq \frac{z}{2R}, \text{ that is } \frac{2(R - \sqrt{R^2 - z})}{z} \geq \frac{1}{R}.$$

We obtain the simplified sufficient condition

$$\forall x \in B \setminus \{0_{d-1}\} : Hf(x) \leq \frac{1}{R}$$

and the proof is concluded. \square

6.3.2. Proof of Theorem 4.4

Proof. We define the test statistic

$$T(X) := \text{dist}(X, \mathcal{C}),$$

and a corresponding test of the form $\varphi(X) = \mathbb{1}_{\{T(X) \geq \tau\}}$.

W.l.o.g. let $\mathbb{O}_d \in \partial\mathcal{C}$ and $\mathcal{C} \subseteq \mathbb{R}^{d-1} \times [0, \infty)$. Now, let $\mu = \mathbb{O}_d$ and $z = (\mathbb{O}_{d-1}, R)$ so that by construction $\mu \in B(z, R) \subseteq \mathcal{C}$.

For $\tau > 0$, we have

$$\begin{aligned} \text{dist}(X, \mathcal{C}) \geq \tau &\implies \text{dist}(X, B(z, R)) \geq \tau \\ &\implies \left\| \frac{1}{\sqrt{n}}\epsilon - z \right\| - R \geq \tau. \end{aligned}$$

Now, using (6.3), we obtain

$$\begin{aligned} \left\| \frac{1}{\sqrt{n}}\epsilon - z \right\| - R &= \sqrt{\left\| \frac{1}{\sqrt{n}}\epsilon - z \right\|^2} - R \\ &= \sqrt{\frac{1}{n}\|\epsilon_{1:(d-1)}\|^2 + \left(\frac{1}{\sqrt{n}}\epsilon_d - R\right)^2} - R \\ &\leq \sqrt{\frac{1}{n}\|\epsilon_{1:(d-1)}\|^2 + \left(R + \frac{1}{\sqrt{n}}|\epsilon_d|\right)^2} - R \\ &\leq R + \frac{1}{\sqrt{n}}|\epsilon_d| + \frac{\|\epsilon_{1:(d-1)}\|^2}{2n(R + \frac{1}{\sqrt{n}}|\epsilon_d|)} - R \\ &\leq \frac{1}{\sqrt{n}}|\epsilon_d| + \frac{\|\epsilon_{1:(d-1)}\|^2}{2nR}, \end{aligned}$$

that is

$$\mathbb{P}_\mu(T(X) \geq \tau) \leq \mathbb{P}\left(\frac{1}{\sqrt{n}}|\epsilon_d| + \frac{\|\epsilon_{1:(d-1)}\|^2}{2nR} \geq \tau\right).$$

Clearly, this bound holds generally in the sense

$$\sup_{\mu \in \partial\mathcal{C}} \mathbb{P}_\mu(\varphi(X) = 1) \leq \mathbb{P}\left(\frac{1}{\sqrt{n}}|\epsilon_d| + \frac{\|\epsilon_{1:(d-1)}\|^2}{2nR} \geq \tau\right).$$

Now, based on the general property

$$\mathbb{P}(A \geq \tau_1) \leq \frac{\alpha}{2} \wedge \mathbb{P}(B \geq \tau_2) \leq \frac{\alpha}{2} \implies \mathbb{P}(A + B \geq \tau_1 + \tau_2) \leq \alpha$$

for random variables A and B and by using (6.2.I) and (6.2.II), we obtain the rejection threshold

$$\tau := \sqrt{\frac{2}{n}v_{\alpha/4}} + \frac{d}{2nR} + \frac{2}{nR}\sqrt{dv_{\alpha/2}} + \frac{1}{nR}v_{\alpha/2} \approx \max\left(\frac{1}{\sqrt{n}}, \frac{d}{nR}\right)$$

for a fixed level $\alpha \in (0, \frac{1}{2})$.

On the other hand, w.l.o.g., choose $\mu = (\mathbb{O}_{d-1}, -\rho)$. Then we have

$$\text{dist}(X, \mathcal{C}) \leq \tau \implies X_d \geq -\tau \iff \epsilon_d \geq \sqrt{n}(\rho - \tau),$$

so that it is sufficient to ensure

$$\sup_{\mu \in \mathcal{A}_\rho} \mathbb{P}_\mu(\varphi(X) = 0) \leq \mathbb{P}(\epsilon_d \geq \sqrt{n}(\delta - \tau)) \leq \beta \in (0, \frac{1}{2}).$$

which leads to the condition

$$\rho \geq \tau + \sqrt{\frac{2}{n}v_\beta} \sim \tau.$$

This concludes the proof. \square

6.3.3. Proof of Theorem 4.5

Proof. This is a variation on the proof of Theorem 3.6. Using the same construction and notation as previously, and taking $d \geq 3$, let now for $i \in \{0, 1\}$

$$\nu_i = \delta_R \otimes \tilde{\nu}_i^{\otimes d-1}.$$

Since the mutual deterministic coordinate $\mu_1 = R$ is irrelevant for the total variation distance between the resulting distributions \mathbb{P}_0 and \mathbb{P}_1 , the bounds in Step 2 of the proof of Theorem 3.6 also hold here with $d - 1$ instead of d .

The most important modification arises when calculating ρ : Now, if at least $\frac{d-1}{4}$ of the coordinates take the value u , computing the Euclidean distance of μ from \mathcal{C} and using (6.3) leads to

$$\begin{aligned} \rho^*(\mathcal{C}) &\geq \sqrt{R^2 + \frac{d-1}{4}u^2} - R \geq \frac{(d-1)u^2}{8\sqrt{R^2 + \frac{d-1}{4}u^2}} \\ &\geq \frac{(d-1)u^2}{8R + 4\sqrt{d-1}u} \\ &\geq \frac{1}{16} \min\left(\frac{(d-1)u^2}{2R}, \sqrt{d-1}u\right) \\ &\sim \min\left(\frac{1}{\ln(d)^4} \cdot \frac{d}{nR}, \frac{1}{\ln(d)^2} \cdot \frac{\sqrt{d}}{\sqrt{n}}\right). \end{aligned}$$

This concludes the proof. \square

6.4. Proofs for Section 5

6.4.1. Proof of Theorem 5.1

Proof. W.l.o.g., let $z = \mathbb{O}_d$. We prove independently that $\rho^*(\mathcal{C})$ is lower and upper bounded by the right hand side of (5.1).

Lower Bound. Let $\nu_0 = \delta_{Re_1}$, giving rise to the density function

$$F_{\nu_0}(x) := \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_1 - R)^2\right) \prod_{i=2}^d \exp\left(-\frac{n}{2}x_i^2\right).$$

On the other hand, for a suitable $h > 0$ specified in a moment, let ν_ρ be the uniform distribution on

$$P_h := \{(R, h \cdot v) \mid v \in \{-1, 1\}^{d-1}\}$$

Since each element of P_h has Euclidean distance $\sqrt{R^2 + (d-1)h^2} - R$ from \mathcal{C} , which should correspond to ρ , we set $h^2 = \frac{(R+\rho)^2 - R^2}{d-1}$. This gives rise to the following density function:

$$\begin{aligned} F_{\nu_\rho}(x) &:= \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_1 - R)^2\right) \frac{1}{2^{d-1}} \sum_{v_2, \dots, v_d \in \{-1, 1\}} \prod_{i=2}^d \exp\left(-\frac{n}{2}(x_i - h \cdot v_i)^2\right) \\ &= \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_1 - R)^2\right) \frac{1}{2^{d-1}} \prod_{i=2}^d \exp\left(-\frac{n}{2}x_i^2 - \frac{n}{2}h^2\right) 2 \cosh(nhx_i) \\ &= \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_1 - R)^2 - (d-1)\frac{n}{2}h^2\right) \prod_{i=2}^d \exp\left(-\frac{n}{2}x_i^2\right) \cosh(nhx_i), \end{aligned}$$

so that

$$F_{\nu_\rho}^2(x) := \left(\frac{n}{2\pi}\right)^d \exp\left(-n(x_1 - R)^2 - (d-1)nh^2\right) \prod_{i=2}^d \exp(-nx_i^2) \cosh^2(nhx_i).$$

Now, using the fact that $\mathbb{E}[\cosh^2(aY)] = \exp(a^2\sigma^2) \cosh(a^2\sigma^2)$ for $Y \sim \mathcal{N}(0, \sigma^2)$, we have

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{F_{\nu_\rho}^2(x)}{F_{\nu_0}(x)} dx &= \int_{\mathbb{R}^d} \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_1 - R)^2 - (d-1)nh^2\right) \\ &\quad \cdot \prod_{i=2}^d \cosh^2(nhx_i) \exp\left(-\frac{n}{2}x_i^2\right) dx \\ &= \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-(d-1)nh^2\right) \int_{\mathbb{R}} \exp\left(-\frac{n}{2}(x_1 - R)^2\right) dx_1 \\ &\quad \cdot \left[\int_{\mathbb{R}} \cosh^2(nhx) \exp\left(-\frac{n}{2}x^2\right) dx \right]^{d-1} \\ &= \exp\left(-(d-1)nh^2\right) \cdot [\exp(nh^2) \cosh(nh^2)]^{d-1} \\ &= \cosh(nh^2)^{d-1}. \end{aligned}$$

Now, by Taylor expansion we obtain the bound

$$nh^2 \leq 1 \implies \cosh(nh^2) \leq 1 + \frac{e}{2}n^2h^4, \quad (6.6)$$

so that

$$\ln(\cosh(nh^2)^{d-1}) \leq (d-1)\frac{e}{2}n^2h^4 = \frac{n^2}{(d-1)}\frac{e}{2}((R+\rho)^2 - R^2)^2, \quad (6.7)$$

whenever

$$nh^2 \leq 1 \text{ i.e. } \rho \leq \sqrt{\frac{d-1}{n}} + R^2 - R. \quad (6.8)$$

Now, provided that (6.8) holds, (6.7) leads to the following sufficient condition for (6.1):

$$\rho \leq \sqrt{\frac{\sqrt{d-1}}{n}s + R^2 - R}, \text{ where } s := \sqrt{\frac{2}{e} \ln(1 + 4(1-\eta)^2)}. \quad (6.9)$$

It is straightforward to see that (6.9) implies (6.8) as long as

$$d \geq 1 + \frac{2}{e} \ln(5) \text{ i.e. } d \geq 3.$$

It remains to investigate (6.9) a little closer. Application of (6.3) now yields the following, defining $t > 0$ via the relation $R^2 = t^2 s \frac{\sqrt{d-1}}{n}$:

$$\begin{aligned} \sqrt{\frac{\sqrt{d-1}}{n}s + R^2} - R &\geq \frac{\sqrt{s}}{2\sqrt{1+t^2}} \frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}} \\ &\approx \frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}} \min(t^{-1}, 1) \\ &\approx \min\left(\frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}}, \frac{(d-1)^{\frac{1}{2}}}{n \cdot R}\right). \end{aligned}$$

If on the one hand $R \lesssim \frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}} \sqrt{s}$, that is $R \leq t \sqrt{s} \frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}}$ for $t > 0$, we have

$$\sqrt{\frac{\sqrt{d-1}}{n}s + R^2} - R \geq \frac{\sqrt{s}}{2\sqrt{1+t^2}} \frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}} \approx \min\left(\frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}}, \frac{(d-1)^{\frac{1}{2}}}{n \cdot R}\right).$$

Analogously, the case $R \gtrsim \frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}} \sqrt{s}$ also yields

$$\sqrt{\frac{\sqrt{d-1}}{n}s + R^2} - R \gtrsim \min\left(\frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}}, \frac{(d-1)^{\frac{1}{2}}}{n \cdot R}\right).$$

Upper Bound. We define the test statistic

$$T(X) := \|X\|^2 - R^2$$

and a corresponding test of the form $\varphi(X) = \mathbb{1}_{\{T(X) \geq \tau\}}$. Clearly, for any $\mu \in \partial C$, we have $n\|X\|^2 \sim \chi_{nR^2}^2(d)$ and for any $\mu \in \mathcal{A}_\rho$ with $\|\mu\| = R + \rho$, we have $n\|X\|^2 \sim \chi_{n(R+\rho)^2}^2(d)$. Therefore, via

$$\sup_{\mu \in \mathcal{C}} \mathbb{P}(\varphi(X) = 1) \leq \mathbb{P}(Z \geq nR^2 + n\tau), \quad Z \sim \chi_{nR^2}^2(d),$$

and using (6.2.II), a rejection threshold τ is readily given by

$$\tau = \frac{d}{n} + 2\sqrt{\left(\frac{d}{n^2} + \frac{2}{n}R^2\right)v_\alpha} + \frac{2}{n}v_\alpha.$$

On the other hand, in order to satisfy a prescribed level $\beta \in (0, \frac{1}{2})$ for the Type II error, that is

$$\sup_{\mu \in \mathcal{A}_\rho} \mathbb{P}_\mu(\varphi(X) = 0) = \mathbb{P}(Z' \leq nR^2 + n\tau) \leq \beta, \quad \text{where } Z' \sim \chi_{n(R+\rho)^2}^2(d),$$

(6.2.III) yields the sufficient condition

$$d + nR^2 + 2\sqrt{(d + 2nR^2)v_\alpha} + 2v_\alpha \leq d + n(R + \rho)^2 - 2\sqrt{(d + 2n(R + \rho)^2)v_\beta}.$$

Using $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ (for $a, b > 0$) and (6.3) respectively, we obtain two different sufficient bounds for ρ :

$$\begin{aligned} \rho &\geq \sqrt{2} \frac{d^{\frac{1}{4}}}{\sqrt{n}} (\sqrt{v_\alpha} + \sqrt{v_\beta}) + \sqrt{\frac{2}{n}} (\sqrt{v_\alpha} + 2\sqrt{v_\beta}); \\ \rho &\geq \frac{\sqrt{d}}{nR + 2\sqrt{nv_\alpha}} (\sqrt{v_\alpha} + \sqrt{v_\beta}) + \sqrt{\frac{2}{n}} (\sqrt{v_\alpha} + 2\sqrt{v_\beta}). \end{aligned}$$

Therefore, as claimed, the upper bound

$$\rho^*(\mathcal{C}) \lesssim \max \left(\frac{1}{\sqrt{n}}, \min \left(\frac{d^{\frac{1}{4}}}{\sqrt{n}}, \frac{\sqrt{d}}{nR} \right) \right)$$

holds.

This concludes the proof. \square

References

- [1] ARIAS-CASTRO, E., AND CASAL, A. R. On estimating the perimeter using the alpha-shape. *arXiv preprint arXiv:1507.00065* (2015).
- [2] BARAUD, Y. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* 8, 5 (2002), 577–606.

- [3] BARAUD, Y., HUET, S., AND LAURENT, B. Testing convex hypotheses on the mean of a gaussian vector. application to testing qualitative hypotheses on a regression function. *The Annals of Statistics* (2005), 214–257.
- [4] BIRGÉ, L. An alternative point of view on lepsi’s method. *Lecture Notes-Monograph Series* (2001), 113–133.
- [5] BULL, A., AND NICKL, R. Adaptive confidence sets in l_2 . *Probability Theory and Related Fields* 156, 3-4 (2013), 889–919.
- [6] CAI, T. T., AND LOW, M. Testing composite hypotheses, hermite polynomials and optimal estimation of a nonsmooth functional. *The Annals of Statistics* 39, 2 (2011), 1012–1041.
- [7] CARPENTIER, A. Testing the regularity of a smooth signal. *Bernoulli* 21, 1 (2015), 465–488.
- [8] COMMINGES, L., AND DALALYAN, A. Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. *Electronic Journal of Statistics* 7 (2013), 146–190.
- [9] GAYRAUD, G., AND POUET, C. Adaptive minimax testing in the discrete regression scheme. *Probability Theory and Related Fields* 133, 4 (2005), 531–558.
- [10] INGSTER, Y. Minimax detection of a signal for l_q^n -balls with l_p^n -balls removed. *WIAS Preprint* (1991).
- [11] INGSTER, Y., AND SUSLINA, I. Minimax detection of a signal for besov bodies and balls. *Problemy Peredachi Informatsii* 34, 1 (1998), 56–68.
- [12] JUDITSKY, A., AND NEMIROVSKI, A. On nonparametric tests of positivity/monotonicity/convexity. *The Annals of Statistics* (2002), 498–527.
- [13] LEPSKI, O., NEMIROVSKI, A., AND SPOKOINY, V. On estimation of the l_r norm of a regression function. *Probability Theory and Related Fields* 113, 2 (1999), 221–253.
- [14] WINKELBAUER, A. Moments and absolute moments of the normal distribution. *arXiv preprint* (2012).